TWO

Polarization and Conformity

Elemental mercury is the only metal that is liquid at room temperature. Aristotle called it "quicksilver," a term that captures its strange beauty. But this particular beauty is also deadly. Exposure to mercury can lead to a host of symptoms: sensations of bugs crawling under the skin, extreme muscle weakness, hair loss, paranoia, mental instability, and, for high exposure levels, death.[1]

The history of mercury use is riddled with such poisonings. Qin Shi Huang, the first emperor of a unified China, is reported to have died in 210 BCE after taking mercury pills that ironically were intended to make him immortal.[2] Isaac Newton sank into paranoia and insanity at the end of his brilliant life—likely a result of his experiments with mercury. (Posthumous hair samples revealed highly elevated levels of it.)[3]

By the end of the twentieth century, the dangers of mercury were well established, and its use was heavily regulated in the United States, Europe, Japan, China, and elsewhere.[4] Mercury poisoning should have been under control.

And yet beginning around 2000, an American physician named Jane Hightower began to notice a distinctive cluster of symptoms in her patients: hair loss, nausea, weakness, brain fog. These are all associated with mercury poisoning—but these patients did not have lifestyles that should have brought them into contact with heavy metals, and so the diagnosis did not occur to her. Until, that is, a colleague heard a story on public radio about a town where locals suffered hair loss and other ailments of mercury poisoning after eating contaminated fish.[5] On a hunch, this colleague ordered a mercury test for one of Hightower's patients.

Sure enough, the patient's mercury levels were elevated.

The patient also ate a lot of fish. Armed with a new hypothesis, that the strange symptoms were linked to mercury and perhaps to fish, Hightower went back to her other mystery patients with a new question. How often did they eat fish? As it turned out, those patients tended to be wealthy and health-conscious and chose to eat fish *very* often—including many fish high on the food chain, such as shark, swordfish, and tuna.

Over the next few years, Hightower systematically recorded her observations and shared her suspicions with colleagues, including some EPA officials who worked on mercury contamination in seafood. Some of the doctors she spoke with began to look for evidence of mercury poisoning in their own patients. Obstetricians in her hospital warned pregnant women off certain fish, since fetal brains are particularly susceptible to the effects of mercury.[6] Some doctor friends quit eating predatory fish. The hospital cafeteria stopped serving canned tuna.

A local news station ran a story on Hightower's suspicions.[7] Then *20/20*, a national television news program, ran a segment on mercury poisoning and fish.[8] Television crews performed tests of the mercury levels in fish at local supermarkets and discovered that some of them, especially shark and swordfish, were well above levels

deemed safe by the US Food and Drug Administration (FDA). Their coverage of Hightower's claims reached a wide audience—and soon more doctors were monitoring their patients for fish-related mercury poisoning, gradually accumulating a larger and larger body of evidence supporting Hightower's hypothesis.

We often associate scientific discovery with lone geniuses—mercurial madman Isaac Newton, Charles Darwin, Albert Einstein—who, in a moment of revelation, conceive of some new theory fully formed. But real discoveries are far more complicated and almost invariably involve many people.[9] Most scientific advances result from the slow accumulation of knowledge in a community. Guesses and observations come from many directions. These insights gradually spread and accumulate, leading to yet more hypotheses and new ideas for how to gather evidence. Only after a long and collaborative process can we say that scientists have achieved a new discovery. Crucial to this process is the network of human interaction linking scientists to one another.

Although Jane Hightower led the effort to link mercury poisoning with overconsumption of contaminated fish, she did not act alone. It was a colleague who first connected hair loss in Hightower's patient with mercury poisoning. It was a contact at the EPA who, upon hearing about her work, shared recent government studies on mercury in fish. Other doctors informed her of patients with similar symptoms, improving her understanding of the syndrome. Hightower's thinking was informed at every step by evidence from outside her own experience.

Conversely, Hightower's insights helped others make even more progress. As soon as she started to gather evidence, her work began influencing the beliefs and behaviors of those around her—obstetricians, other clinicians, medical associations—who went on to find more evidence and further links. Ultimately, the discovery of a new link between mercury poisoning and seafood consumption occurred

when a community, or *network*, of scientists and doctors, all sharing ideas and evidence, adopted a new consensus.

In this way, those responsible for scientific discovery are bolstered by those around them. Bolstered—but also, sometimes, stymied. Hightower's evidence did not convince everyone she shared it with. To many colleagues, she seemed like an activist with some kind of environmentalist axe to grind, or perhaps just a quack. In fact, there seemed to be good reasons to think that the symptoms Hightower observed could not be from mercury.

In the early 2000s, it was already widely known that some fish contained mercury. Coal-fired power plants emitted a form of inorganic mercury into the air, where it would gradually fall back to earth, mix into ocean water, and be ingested by microbes, which converted it to highly toxic methylmercury. These microbes would then be consumed by small fish, which would be consumed by larger fish, and so on up the food chain. Methylmercury tends to accumulate in animal tissue, so large fish were building up high levels of the toxin. This was why the FDA already had guidelines regulating the level of methylmercury in fish sold commercially—levels that, it turned out, were exceeded by some supermarket supplies.

So the idea that fish contained toxic mercury was not controversial. But precisely because the whole process seemed well-understood, regulators, including the FDA, thought they knew what the dangers were. When presented with Hightower's work, the FDA responded that no one was actually eating enough fish to be poisoned. Many of her colleagues seemed to agree.

Still, Hightower pushed forward with a year-long survey documenting the fish intake, symptoms, and blood mercury levels of a group of patients. She published these results and shared them with a contact at the EPA, who invited her to present her work at a meeting of mercury experts. At the suggestion of another colleague she wrote a resolution about the dangers of methylmercury and how to

tackle them, which was passed by the California Medical Association and San Francisco Medical Society.

With time and ever more evidence, she gradually convinced more and more of her colleagues. Today, government agencies around the world are more savvy about the risks of methylmercury poisoning from fish and have issued guidelines to better control exposure.

On February 28, 1953, around lunchtime, the English biologist Francis Crick called for the attention of his fellow diners at the Eagle Pub, in Cambridge, UK.[10] He had an important announcement to make: he and an American geneticist named James Watson had "discovered the secret of life." That secret, according to Watson and Crick, was the physical structure of a complex molecule, DNA, that contains the basic genetic material for virtually all life on earth.

On the road to discovering the structure of DNA, Watson and Crick drew on many tools.[11] Perhaps the most iconic of these was a set of glorified Tinkertoys they used to represent various atoms and the electrical bonds between them.[12] These building blocks allowed Watson and Crick to test hypotheses about the feasibility of diverse molecular structures.

In most ways, the structures they built were nothing like molecules. The pieces were hundreds of millions times bigger than atoms, and they were painted various colors, which atoms decidedly are not. Electron structure was represented by sticks poking out of balls at different angles. And yet, by experimenting with these blocks, Watson and Crick managed to extract crucial insights into the real structure of DNA.

The kind of reasoning Watson and Crick did with their building blocks is ubiquitous in the sciences. They built a *model* as an aid to understanding and inference. Models can take many different forms:

physical structures developed in labs, computer programs, mathematical constructions of various sorts. Usually, a model is some sort of simplified or otherwise tractable system that scientists can manipulate and intervene on, to better learn about a messier or more complex system that we ultimately care about.[13] Watson and Crick could not play with the actual structures of molecules, but they could manipulate their building blocks instead and use the resulting structures to learn about the real system.

In fact, we introduced an example of this kind of model in the last chapter—though we did not explicitly label it as such. Bayes' rule, remember, is a formula for how people ought to change their beliefs in light of new evidence. To apply Bayes' rule, we first need to think of our confidence concerning our various beliefs as represented by probabilities—basically, numbers between 0 and 1 that have to satisfy some further conditions. This whole picture, where degrees of belief are numbers that can change via Bayes' rule as we collect evidence, can be thought of as a simplified mathematical model of how humans might really change their minds.

Of course, this model will not capture most real cases of inference perfectly. But it can nonetheless provide insights into what is going on when our beliefs evolve as we learn about the world. It captures the idea that beliefs come in degrees, and it sets out conditions under which those beliefs should change. For instance, if the evidence we have is very likely to occur if our belief is true, we should become more confident in that belief. If our evidence is unlikely when the belief if true, we should become less confident. As we argued in Chapter 1, this insight alone is useful for thinking about issues regarding whether science can ever deliver certainty about anything—and whether we should care.

Bayesian belief updating gives us a model of how *individual* beliefs change. But as we have just seen in the case of methylmercury, science often needs to be understood on the level of a community,

not an individual. How do groups of scientists—such as the one Jane Hightower was part of—share knowledge, evidence, and belief? How do they reach consensus? What do these processes tell us about science?

These questions, too, can be studied by developing and examining models. There are many ways to do this, but here, to keep things simple, we focus on just one framework.[14] Where there are other important models to discuss, we do so in the endnotes.

The framework we focus on was introduced in 1998 by economists Venkatesh Bala and Sanjeev Goyal. It is a mathematical model in which individuals learn about their world both by observing it and by listening to their neighbors. About a decade after Bala and Goyal introduced their model, the philosopher of science Kevin Zollman, now at Carnegie Mellon University, used it to represent scientists and their networks of interaction.[15] We use the model, and variations based on it, much as Zollman did.

Why might models be useful here? Communities of scientists are vastly complex. We can investigate them using experiments and case studies, but there are some things that even these powerful methods cannot do for us. For example, we could never track the full progress of an idea, such as that methylmercury was poisoning fish eaters, through an entire scientific network. Where did each scientist first hear of it? When did he or she become convinced it was correct? Who did that scientist share it with? This is especially true of scientific insights that happened in the deep past, and ones that involved large networks of researchers. Models can help fill the gaps in our understanding of how beliefs spread in communities of scientists, and knowledge seekers more generally.

Of course, a model of scientists gathering evidence and communicating with one another cannot capture every detail of how scientific ideas develop and spread. For example, we will not attempt to model the "Eureka moment"—the dawning of that brilliant idea

that moves a field forward. (Though, again, we are skeptical that such moments play the significant role history tends to grant them.) Nor will we model power dynamics between scientists, or the role that prestige and timing play in the uptake of scientific ideas.[16] We focus just on the dynamics of belief and evidence.

Even this very simplified model can give us surprising information that we could get no other way. It provides a new way of thinking about how beliefs spread in a community—and a way to ask how those dynamics would change under various conditions.

The basic setup of Bala and Goyal's model is that there is a group of simple *agents*—highly idealized representations of scientists, or knowledge seekers—who are trying to choose between two actions and who use information gathered by themselves and by others to make this choice. The two actions are assumed to differ in how likely they are to yield a desired outcome. This could represent the choice between eating fish or not and so increasing or decreasing one's risk of mercury poisoning; or it could be regulating smokestack emissions and so increasing or decreasing the risks of acid rain. For a very simple example, imagine someone faced with two slot machines, trying to figure out which one pays out more often.[17]

Over a series of rounds, each scientist in the model chooses one action or the other. They make their choices on the basis of what they currently believe about the problem, and they record the results of their actions. To begin with, the scientists are not sure about which action is more likely to yield the desired outcome. But as they make their choices, they gradually see what sorts of outcomes each action yields. These outcomes are the evidence they use to update their beliefs. Importantly, each scientist develops beliefs based not only on the outcomes of their own actions, but also on those of their colleagues and friends.

A clinician like Hightower, for example, might observe what happens to her own patients and also hear about her colleagues' patients. She will use all of these observations in deciding whether she thinks her patients' symptoms are due to mercury poisoning. Similarly, while at the casino you might favor one slot machine, but after hearing from all your friends that they hit jackpots on another, you might change your mind.

In the model, one of the two actions—call it action B—is, in fact, better than action A. (To keep this straight, remember that A is for "All right," but B is for "Better.") But figuring out which action is superior is not necessarily easy. A crucial assumption in this model is that evidence is probabilistic, meaning that when the scientists investigate the world—test a slot machine or warn a sick patient off fish—the results are not always the same. Action B is better than action A because, on average, it yields better results. But there can be many individual instances when action A happens to yield a better result.

In this way, we can think of action B as similar to a biased coin. It may land heads up more often than an ordinary coin—but that does not mean that it never lands tails up. And if you flip a biased coin and an unbiased coin some number of times, there is no guarantee that the biased one will land heads up more often. It is merely *likely* that it will do so.

Not all science looks like this. If you were investigating the laws of gravity, for instance, and you dropped a bowling ball off the top of the Empire State Building again and again, very carefully timing it on each attempt, the results would be remarkably consistent. Likewise for mixing natural gas, oxygen, and a flame: we know what will happen.

But in many types of science, evidence is not so dependable. Again, think of methylmercury. Individual sensitivity to the toxin varies widely, meaning that two people eating the same amounts of

swordfish might show very different symptoms. To make matters worse, the symptoms take time to develop. In retrospect, it is easy to look back on Chinese emperors taking mercury tablets to become immortal and think, "How stupid! How did they miss that the stuff is toxic?" But mercury has historically been used again and again in medicine, because without statistical methods it is actually quite difficult to definitively link its use to its harms. The effects are too variable. In cases like this, scientific consensus is hard to reach, and models like the one we are describing can help us understand how that consensus comes about.

We should also emphasize that, although our examples come from science and we are calling the agents in our model "scientists," these models can represent any group of people who are trying to make their way in an unpredictable world. All of us act as scientists sometimes, when we make decisions based on our own experiences and those of our friends. Ever buy a car? There is a good chance that you took it for a test drive and asked the dealer some questions. You were gathering evidence before making a decision. Did you also ask your friends or relatives for advice? Or look at online reviews? If so, you consulted a network of other agents who likewise had gathered evidence, and you used their experiences to influence your beliefs—and ultimately your actions. So these models can apply very broadly. (We will return to this point in Chapter 4.)

We described Bala and Goyal's models as mathematical. At this point, you might wonder where the math comes in. Let us dive into the details a bit more. We have been using anthropomorphic language, talking about "scientists" who "decide" to "act" on the basis of their "beliefs." But in fact we are talking about computer simulations—there are no real decisions here, no physical actions, and no minds that could hold beliefs. Instead, we have an abstract network consisting of a collection of "nodes," each of which may or may not be connected to other nodes by what is called an "edge."

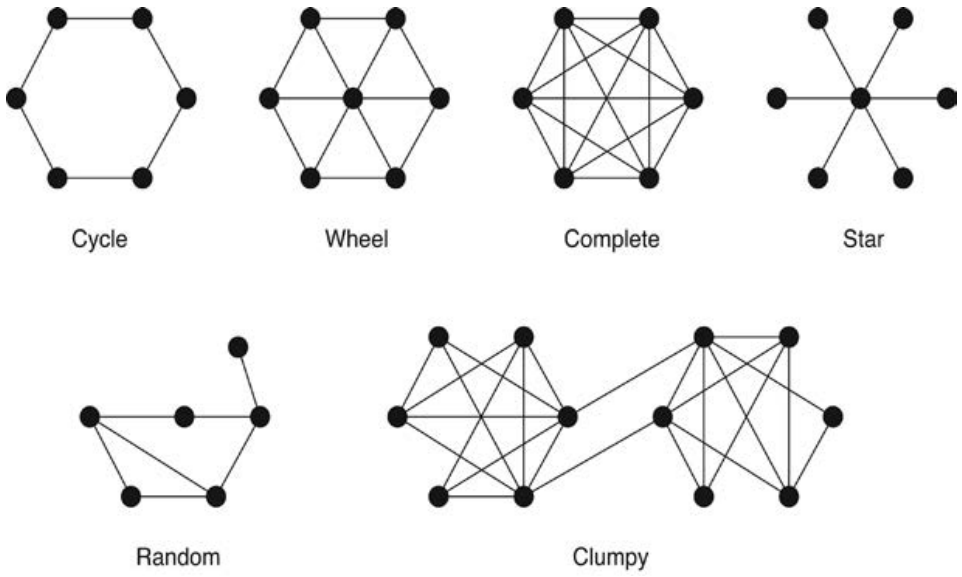

Figure 1. A collection of communication networks. In each network, the nodes represent individuals, or agents, and the connections between them, called edges, represent social ties. Some networks, like the complete, are more densely connected, and others, like the cycle, are more sparse. The clumpy network involves cliques. In the star and wheel networks, some individuals are more central than others. These structures influence how beliefs spread through the network.

Each node represents a scientist, and each edge connects two scientists who have access to each other's results.

These networks can take different shapes. Figure 1 shows some examples of what the communication networks of scientists might look like. Some of these follow patterns: the cycle is a ring with each individual connected to two others; the complete network directly connects all agents to all other agents; and both the star and the wheel have one central node, with the rest of the individuals in the wheel loosely connected and in the star not connected at all. Real human networks are not so neat. They often have substructures that mimic the more regular ones, but they are also "clumpy," with random links between well-connected cliques.[18] As we will see,

these structures are often important in determining how information and ideas flow through a group.

How does a node make decisions or take actions? In the model, each node—each scientist—is associated with a number between 0 and 1. This number represents the scientist's level of certainty, or credence, that action B is better than action A. An assignment of .7 would mean that particular scientist thinks there is a 70 percent chance that action B is better than action A. Which action the scientist takes is wholly determined by this number. If it is greater than .5, the scientist performs action B—by which we mean that we simulate pulling a slot machine some number of times and counting the number of times it pays off. Then we use Bayes' rule to update the scientist's credence in light of this result, and likewise update the credences of all of the other neighboring scientists on the network.

If the scientist's belief is less than .5, he or she performs action A. In the simplest version of the model, we assume that everyone knows that this action works exactly half the time.[19] You can think of this as a situation in which, say, a new medical treatment (action B) has been introduced to a market where another well-studied and well-understood treatment (action A) is already available.[20] Doctors are interested only in whether the new treatment is better than the old one; they already know how well the old one works. The fact that we have a network of scientists, however, means that any particular scientist can get evidence of the new treatment's efficacy from their neighbors, even if they do not perform that action themselves. This is like the other physicians who learned of Dr. Hightower's results, even though it never occurred to them to test their own patients' mercury levels.

Figure 2 shows an example of what this process might look like. First, in (a) we see a network of six nodes (scientists) and edges (their connections). Each scientist has a credence ranging from 0 to

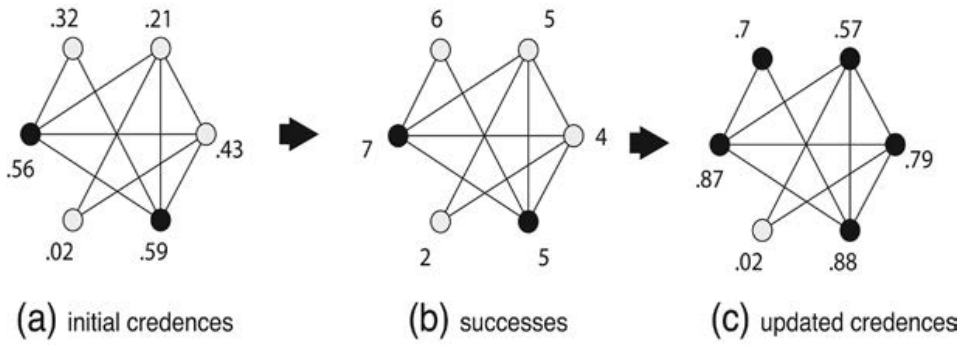

Figure 2. An example of updating and experimentation in a Bala and Goyal–style model. Scientists start with initial credences (a) and use these to decide how they will test the world (b). Light nodes represent those taking action A, and dark nodes, B. In (c) we see that scientists who observe tests of action B update their credences.

1 that action B is better. We can also see that on the basis of their credences in this particular network, four scientists will perform A (the light nodes) and two will perform B (the dark ones). Say they each perform their action ten times. In (b) we can see an example of results they might have obtained (2, 5, 7, etc.) Then in (c) we see how each scientist changes credences using Bayes' rule on the basis of the outcomes observed by themselves and their neighbors. Anyone connected to someone who tried action B—the new, unknown treatment—will update their beliefs. (The scientist with credence .02 does not update since that scientist is not connected to anyone trying action B.) In this case all but one scientist increased their confidence in B, since, as expected, it tended to succeed more often than A. In fact, we can see that when they act next, five scientists will try B instead of A.[21]

This process continues stepwise (try actions, update credences, try actions, update credences) until the scientists have converged on a consensus. This can happen if all of the scientists have sufficiently high credence—greater than .99—that action B is better; or all of

them have sufficiently low credence, less than .5, so that no one in the network ever performs action B, in which case they will not learn anything further about it. In the first case, we say the network has converged to the true belief. In the second, we say it has converged to the false one. In general, these models tend to converge to the true consensus—that is, the whole network comes to believe that action B is better. But, as we will see, they sometimes go to the false one.[22]

What we want to understand is this: Under what circumstances do networks of scientists converge to false beliefs?

Stomach ulcers are painful sores in the lining of the stomach. It turns out that they are caused by a kind of bacteria known as *H. pylori*.[23] Decisively showing that bacteria cause ulcers ultimately earned the 2005 Nobel Prize for two Australian medical researchers, Robin Warren and Barry Marshall, who managed to convince their fellow scientists of this relationship during the 1980s. But it is a bit strange to say that Warren and Marshall *discovered* the link. In fact, the theory that ulcers were caused by bacteria dates back to 1874, when a German bacteriologist by the name of Böttcher and a French collaborator, Letulle, isolated bacterial colonies in an ulcer and argued that the bacteria were the ulcer's cause.[24] During the following decades, evidence slowly accumulated that bacteria were, indeed, responsible for ulcers.

But the bacterial theory was not the only one available. The other possibility, also accepted by many doctors and scientists, was that stomach acid was the culprit. In the early twentieth century, scientists investigated both theories and found evidence in favor of each. But then, in 1954, the bacterial theory suffered a devastating setback. Gastroenterologist E. D. Palmer biopsied the stomachs of more than one thousand patients and found no evidence of bacteria

at all.[25] The conclusion seemed to be that bacteria could not live in the human stomach, meaning that they could not possibly cause ulcers.

Palmer's results essentially ended attempts to confirm the bacterial theory—aside from a few isolated doctors who continued to successfully treat ulcer patients with antibiotics. (Inhibiting gastric acid also helped—though ulcers treated in this way tended to return.) It was not until almost thirty years after Palmer published his results, when Warren observed a new strain of bacteria in stomach biopsies taken near tissue with ulcers, that serious research on the bacterial theory picked up steam again. Later, Marshall managed to isolate and cultivate the new strain, showing definitively that bacteria *could* live in the human stomach after all.

Even with these strong results, Warren and Marshall faced significant skepticism. The acid theory was widely held and deeply ingrained. The resistance was so strong that Marshall resorted to dramatic stunts to attract attention—and adherents—to their theory. In a fit of pique, he apparently drank a petri dish full of *H. pylori* himself and then successfully treated the ensuing ulcer with antibiotics.[26] Ultimately, Warren and Marshall managed to persuade their colleagues that the bacterial theory was right. But this episode could very well have gone differently. Had there not been a few scientists willing to give the bacterial theory a chance, we might still be using antacids to treat recurring ulcers.

How could this happen? One of the most startling findings from the Bala-Goyal models is just how strongly people's beliefs can influence one another. If we imagined a group of agents with no network connections gathering probabilistic evidence (and not sharing it), we would expect some of them to end up with the right theory and some with the wrong one. For instance, scientists who play the better slot machine and happen to lose all their money may give up on that machine for good. But with no communication, we should

not expect much correlation between the various scientists' beliefs. Some scientists would have good luck and stick with the better machine; others would not. Each conclusion would be independent from all the others.

Once scientists start to share evidence, however, it becomes extremely likely that they will all come to believe the same thing, for better or worse.[27] Notice that this happens in the models only because the scientists share evidence. There is no psychology here. No one is imitating anyone else, no one is trying to conform, no one is smarter or dumber than the others. There are no thought leaders or sheeple.

Why does it happen? Imagine a group of scientists gathering and sharing data. Suppose a few of them try the better action—reducing fish consumption, say, on the hypothesis that eating too much fish can cause mercury poisoning. As they continue to gather evidence, it starts to influence their colleagues and neighbors, just as we saw in the Hightower case. Some of these come to believe the right theory and now start to gather evidence about it themselves. They, in turn, can persuade new colleagues and neighbors. The belief spreads throughout the network until everyone agrees.

Notably, this means that a successful *new* belief can spread in a way that would not have been very likely without the ability to share evidence. Suppose that almost every scientist starts with an extant belief (say, the mercury in fish is not poisoning people). We do not expect them to gather evidence about mercury and fish—why would they? Without data sharing, the chance that each independently decides to test this new possibility is miniscule. With data sharing, however, it takes just one scientist to start testing a new hypothesis for it to start catching hold throughout the scientific network (if the scientist gets positive results).

Figure 3 shows what this might look like. It is a simplified image (showing just the updating of credences, but not the successes) of

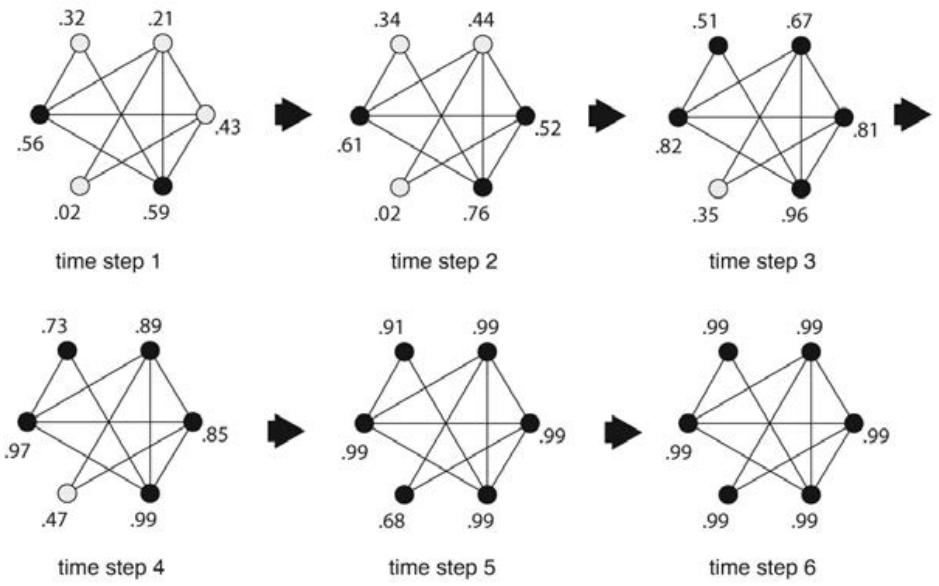

Figure 3. An example of a network that achieves convergence on true beliefs. Light nodes represent belief in A and dark nodes belief in B. In each time step agents are testing their beliefs and updating their credences on the basis of their results and their neighbors' results. As time goes on, more agents have high credences in the true belief until the entire network becomes essentially certain that action B is better.

the process like that shown in figure 2. In each subsequent round, more scientists are persuaded by the results of their neighbors to try the better action, and eventually it spreads throughout the network.

This is the optimistic outcome. As we argued in the Introduction, the social spread of knowledge is a double-edged sword. It gives us remarkable capabilities, as a species, to develop sophisticated knowledge about the world, but it also opens the door to the spread of false belief. We see this in the models as well: especially when scientists tackle hard problems, they can all come to agree on the wrong thing. This happens when a few scientists get a string of misleading results and share them with their colleagues. Scientists who might have been on track to believe the true thing can be derailed by their

peers' misleading evidence. When this happens, the scientists would have been better off *not* getting input from others.

It is worth taking a moment to let this sink in. Usually, when scientists behave rationally but gather uncertain data, sharing evidence helps the whole group get to the right belief, even persuading those who were initially skeptical. But sometimes this process backfires, and communication between scientists actually leads to a consensus around the false belief. Remember the Vegetable Lamb. Without communication among learned scholars, this bizarre belief would never have gone anywhere. The sharing of evidence ("I tasted its wondirfulle flesh!") convinced many with correct beliefs that the wrong thing was true.

This trade-off, where connections propagate true beliefs but also open channels for the spread of misleading evidence, means that sometimes it is actually better for a group of scientists to communicate less, especially when they work on a hard problem. This phenomenon, in which scientists improve their beliefs by failing to communicate, is known as the "Zollman effect," after Kevin Zollman, who discovered it.[28] If everybody shares evidence, a chance string of bad data can persuade the entire group to abandon the correct theory. But in a group where not everyone listens to everyone else, pockets of scientists can be protected from misleading data and continue to gather evidence on the true belief that eventually persuades the rest of the community.[29]

Another way to put this is that some temporary diversity of beliefs is crucial for a scientific community. If everyone starts out believing the same thing, they can fail to try out better options. It is important for at least a few people to test different possibilities so that the group will eventually find the best one. One way to maintain this diversity of beliefs for a long enough time is to limit communication, so that researchers' beliefs do not influence one another too much while they test different theories.[30]

As Zollman himself points out, the Zollman effect can help explain how Palmer's results finding no bacteria in the stomach had such a dramatic effect—and why the medical establishment held fast to a false theory for so long.[31] Physicians were tightly connected to one another, so a single result—even though it turned out to be misleading—convinced nearly all of the gastroenterologists in the world that they should abandon what turned out to be the true theory of ulcers. Taking the actions they did was very likely the rational thing to do given Palmer's evidence, which seems to have been very strong. But the structure of the community meant that rational actions by every individual actually made the false belief persist. Had fewer scientists known about Palmer's results, the bacterial theory might have won out sooner.

Of course, Warren and Marshall did, eventually, return to the bacterial theory. If we add to the model the fact that scientists sometimes test the alternative theory—they sporadically or accidentally perform action B, even though they generally do not expect it to be better—they can overcome the Zollman effect, much as they would if they were less tightly connected with one another. But it can be a slow process and relies on luck. On the other hand, it works precisely because of evidence sharing: if strong evidence for a surprising new theory appears in this random way, the connections between scientists will allow the better theory to eventually take hold and spread.[32]

Polly Murray was suffering from fatigue, terrible headaches, and joint pain so severe that she struggled to move.[33] She had seen doctors, but none of them had managed to help her. Many, in fact, hinted that her symptoms might by psychosomatic—or to put it more bluntly, they thought she was a nut. But as Murray meticulously documented, she was not the only one with these symptoms.

Many of her friends and their children, all living in the small town of Lyme, Connecticut, suffered from the same strange cluster of ailments. Two of her children had been diagnosed with juvenile rheumatoid arthritis. This is a rare disease and it is not infectious—it seemed exceedingly unlikely that there could be an epidemic of it.

In 1975, Connecticut health officials took Murray's case to Allen Steere, a rheumatologist working on a fellowship at Yale. Steere met with Murray, and she showed him her list of neighbors with the same symptoms.[34]

Steere's extensive investigation into the possible causes of the ailment eventually yielded a diagnosis: a new tick-borne illness later named Lyme disease, after the town where Murray and her friends lived.[35] A few years later, the strain of bacteria responsible was isolated and named *Borrelia burgdorferi* (after Willy Burgdorfer, who did the isolating).[36] This discovery had a massive impact on patients like Polly Murray. After treatment with antibiotics, many of them regained lives previously lost to debilitating pain. September 24 was declared "Allen Steere Day" in Connecticut to celebrate his findings.

Fast forward twenty-five years. Allen Steere was receiving death threats and hate mail from Lyme patients across the country. Security guards had to be hired to protect him at public appearances. The New England Medical Center, where he was now chief of rheumatology, employed an expert who spent hours each week monitoring the public threat to his safety.

What had happened?

Lyme disease is caused by a spirochete—a type of bacteria shaped like a spiral or helix, like those that cause syphilis. And like syphilis, the disease proceeds in stages. Initial infection causes flulike symptoms: fever, headache, joint aches, and often, but not always, a distinctive rash in the shape of a bull's-eye.[37] As the spirochetes spread throughout the body, some patients develop more alarming

symptoms: meningitis, encephalitis, facial paralysis, and mental disturbances.[38]

As with any infection, the human immune system responds by attacking the invader, producing antibodies that help it identify and root out the Lyme spirochete. In many cases, though, this is not enough to totally suppress the infection. *Borrelia* uses its distinctive shape to wriggle into tissues throughout the body, and it employs a host of nasty tricks to hide from the immune system. When left untreated, late-stage Lyme causes the sorts of symptoms that first brought Polly Murray to see Allen Steere: crippling joint pain, numbness and pain in the extremities, brain fog, insomnia, extreme fatigue, and maladies such as serious cognitive impairments.[39]

This much, at least, is relatively uncontroversial. But what happens after Lyme is treated by antibiotics? This question is at the heart of what has become known as the "Lyme wars." It is the Lyme wars that put Allen Steere's safety at risk.

On one side are those who hold the view, widespread within the medical establishment and endorsed by groups such as the US Centers for Disease Control and Prevention (CDC), that a single round of treatment with antibiotics is generally enough to eliminate the Lyme spirochete, and so to cure a patient of the disease.[40] On the other side are a large number of Lyme disease patients who have already undergone antibiotic treatment but who continue to experience debilitating symptoms typical of the disease. On the basis of their experiences, some "Lyme-literate" doctors have developed treatment programs for "chronic Lyme disease," usually involving repeated rounds of heavy antibiotic use.

In the early 1990s, observing the emergence of the Lyme-literate doctor movement, Steere grew concerned that the diagnosis of Lyme disease had become a catchall for other diseases such as fibromyalgia and chronic fatigue syndrome (themselves both poorly understood and controversial). After investigating patients referred

to him for Lyme, he formed the opinion that many did not have the disease. Knowing that long-term antibiotic use has serious side effects, he began to publicly advocate for greater caution in Lyme diagnosis and treatment.

Thus began a decades-long battle (which is still raging) over chronic Lyme disease. Steere, and most professional doctors' groups and disease control centers, contend that chronic Lyme is actually a combination of other diseases, plus, perhaps, a mysterious post-Lyme syndrome that might involve a continued immune response to Lyme after it has already been treated. They argue that long-term antibiotic treatments do serious harm to sick patients, without any benefits.[41] Most chronic Lyme patients, they point out, do not test positive for the Lyme spirochete, and four large studies conducted by the National Institutes of Health have each shown that long-term antibiotic treatments do not improve these patients' symptoms.[42]

On the other side of the debate are the patients, Lyme-literate physicians, and various advocacy organizations. They contend that Lyme spirochetes often hide in the body, avoiding total eradication by standard antibiotic treatments, and that long-term antibiotics are an effective treatment.[43] The doctors involved claim to have successfully treated thousands of patients. They refer to evidence showing that Lyme can survive aggressive antibiotic treatment in dogs, mice, and monkeys[44] and can subsequently reinfect ticks and other hosts with live spirochetes despite sometimes failing to show up in standard tests.[45] Disaffected by what they see as a wall of opposition from mainstream researchers, the Lyme Disease Foundation even started its own publication—the *Journal of Spirochetal and Tick-Borne Diseases*—to publish results defending the existence of chronic Lyme.

At stake in this debate is the well-being of thousands of suffering patients. They vilify Steere and others who maintain that those with chronic Lyme should not be treated indefinitely. Some argue that

these doctors are conspiring to hide the truth about chronic Lyme—possibly because they are in cahoots with insurance companies that do not want to pay for long-term treatment.[46]

If threats of violence against doctors seem extreme, note that the medical establishment has its own weapons. Patients are denied insurance coverage for expensive treatments that they claim reduce their symptoms. Doctors willing to prescribe long-term antibiotic treatments are often regarded as quacks and pariahs by their colleagues and by medical licensing boards. Some of the most prominent of these doctors, such as the beloved Charles Ray Jones, who has treated thousands of children for chronic Lyme, have been disciplined by licensing boards or had their licenses suspended.[47]

On both sides, the Lyme wars have extended far beyond discussions over coffee at academic conferences and in the pages of medical journals. And one side is putting people's lives at risk. The only question is which.

On June 14, 2017, in Alexandria, Virginia, a group of Republican congressional representatives met to practice for the Congressional Baseball Game for Charity, which was scheduled for the following day.[48] Suddenly, mid-practice, shots rang out from near the third-base dugout. Congressman Steve Scalise was hit in the hip; a lobbyist, a congressional aide, and a police officer assigned to protect Scalise were also shot and injured. The gunman was shot and died of his wounds.

The shots were fired by a left-wing extremist named James Thomas Hodgkinson. Hodgkinson reportedly belonged to Facebook groups with names like "The Road to Hell Is Paved with Republicans," where he posted vitriolic anti-Trump comments daily.[49]

Two months later, white supremacists, neo-Nazis, and other nationalist and nativist extremists marched through Charlottesville,

Virginia, carrying torches and chanting anti-Semitic, racist, and pro-Trump slogans.[50] Violence broke out between the "Unite the Right" crowd and counterprotesters, injuring fourteen people. The following day, a twenty-year-old white supremacist named James Alex Fields Jr. drove his car into a counterprotest. He injured nineteen people and killed a thirty-two-year-old woman named Heather Heyer. In the month before the attack he reportedly had posted photos of Nazis, swastikas, and pro-Trump memes on his Facebook page, as well as pictures of alt-right icons such as Pepe the Frog.[51]

The term "polarization" originated in physics to describe the way some electromagnetic waves propagate in two oppositely oriented ways. By the mid-nineteenth century, political pundits had embraced this metaphor, of two opposite ways of being, to describe disagreements in a state dominated by two parties. Today it captures the broad sense that Democrats and Republicans, Labour and Tories, left-wing and right-wing, are increasingly divided in their beliefs and moral stances.

Hallmarks of polarization include individuals on two sides of an issue who tend to move farther from consensus, rather than closer to it, as debate progresses. In some instances of political polarization, moral mistrust breeds between those who disagree, sometimes leading to violence, as in the shooting of Steve Scalise and the killing of Heather Heyer.

In the case of chronic Lyme disease, we see a situation where a *scientific* community has polarized over a set of scientific beliefs in much the way that some communities polarize over political beliefs. Here, too, the situation has progressed to threats of violence.

This situation may seem surprising. We tend to think of political stances and scientific beliefs as importantly different. Political stances are motivated by social values: moral norms, religious beliefs, and beliefs about social and economic justice. We adopt political positions because we want to promote something we value in

our country and our lives. Scientific beliefs, on the other hand, are supposed to be value-free (arguments from Chapter 1 notwithstanding). In an ideal science, thinkers adopt beliefs that are supported by evidence, regardless of their social consequences.

In fact, this is not how science works. Scientists are people; like anyone else, they care about their communities, their friends, and their country. They have religious and political beliefs. They value their jobs, their economic standing, and their professional status. And these values come into play in determining which beliefs they support and which theories they adopt.[52]

That said, it is not clear, in the case of the polarization over Lyme disease, that differing values play much of a role. The physicians on both sides of the debate seem to have the same values. Allen Steere has devoted his professional life to studying and treating the disease. His objections to patients taking heavy doses of antibiotics seem to be genuinely motivated by concern for their health and safety. At the same time, doctors such as Charles Ray Jones are trying to treat patients who are truly suffering, and, on their own reports, they are succeeding in doing so. Everybody involved wants to protect and cure the afflicted.[53]

Besides having the same values, the two sides in the chronic Lyme case have access, for the most part, to the same evidence. They can, and often do, read the same journal articles about Lyme disease. They see patients with similar symptoms. Inasmuch as Lyme-literate physicians prescribe long-term antibiotics and most other physicians do not, these groups will not always observe patients undergoing the same sorts of treatments, but all of them read the same reports of randomized controlled trials on the effects of antibiotic treatments, and they can discuss other doctors' clinical observations.

So how have things gotten so polarized? The models of scientific networks we have described in this chapter suggest that scientific

communities should tend strongly toward consensus as they gather and share evidence. Eventually, influence and data flowing between researchers should sway the whole group one way or another.

Or perhaps not. The models we have considered so far assume that all scientists treat all evidence the same way, irrespective of the source. But is that reasonable? Do all scientists trust one another equally? Do they consider all other researchers equally reliable?

Consider a small alteration to the model we introduced earlier. Suppose scientists in a network do not treat all the evidence the same way but instead take into account how much they trust the colleague who is sharing research with them. This is hardly an unreasonable thing to do. It is, in fact, an essential part of science—and scientific training—to evaluate the quality of the evidence one encounters, and to exercise judgment in reacting to putative evidence. Taking into account the source of reported data is surely a natural way to do this. Scientists who rely on studies written by known quacks are arguably abdicating their responsibilities.

How can we include this sort of "trust" in the Bala-Goyal model? Here is one suggestion. Suppose scientists tend to place greater trust in colleagues who have reached the same conclusions they have reached, and less in those who hold radically different beliefs. Again, this is not so unreasonable. We all tend to think we are good at evaluating evidence; it is only reasonable to think that those investigating similar problems, who have reached different conclusions, must not be doing it very well.[54]

We can thus change how the scientists in our model update their beliefs in light of new evidence. The rule we have used so far, Bayes' rule, takes for granted that we are certain that the evidence we are considering was really observed: there were no errors, no subterfuge, no miscommunications. This is a highly idealized case. Usually, when we encounter evidence, it is not perfectly certain. In such cases, there is a different rule that can be used to update your be-

liefs, called "Jeffrey's rule," after Princeton philosopher Dick Jeffrey, who proposed it. Jeffrey's rule takes into account an agent's degree of uncertainty about some piece of evidence when determining what the agent's new credence should be.[55]

But how much uncertainty should the scientists assign to any particular piece of evidence? Suppose they do this by looking at how far the other scientists' beliefs are from their own, and letting that distance determine their degree of uncertainty. Reading Allen Steere's newest article, a Lyme-literate physician does not fully trust the reported results. Hearing about the clinical experiences of Charles Ray Jones, an establishment researcher is skeptical. In one version of this model, the scientists simply stop listening at some point and do not update their beliefs at all on the basis of evidence produced by someone who disagrees with them too much. In another version, the scientists could think that the scientists who disagree too much are corrupt or otherwise trying to mislead them and therefore assume that the evidence they have shared is actively fabricated. In this case, they would update their beliefs in the other direction.[56]

This small change to the model radically alters the outcomes. Now, instead of steadily trending toward a consensus, either right or wrong, scientists regularly split into polarized groups holding different beliefs, with each side trusting the evidence of only those who already agree with them.[57] Initially, scientists' beliefs are randomly distributed throughout the network. Most scientists begin by listening to, and updating on the basis of, the evidence produced by most other scientists. But over time, groups of scientists begin to pull apart until eventually you have two groups with opposite beliefs who do not listen to each other at all.

Such a model does not capture the moral anger we see in the case of chronic Lyme, or in political polarization. But we do see that under fairly minimal assumptions, entire scientific communities can

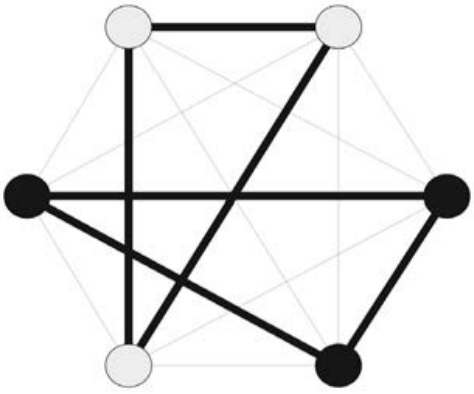

Figure 4. A complete, that is, fully connected, network in which agents are polarized in that they have stable, opposing beliefs. Light nodes represent those taking action A, and dark nodes, B. The weights of the connections between the nodes represent trust between agents—which translates into belief that other agents share real data. Within each group, agents trust others' data, but they do not trust data from the other group.

split into two groups with opposite beliefs. Even worse, this sort of polarization is stable: no amount of evidence from the scientists who have adopted the correct belief will be enough to convince those who adopted the wrong belief. And the polarization does not depend on individuals not seeing the evidence of those with different beliefs. They receive this evidence just as before. They simply do not believe it.

Figure 4 shows a network in which all people see each other's evidence (a complete network) but that has moved toward polarization. The shade of the nodes represents which belief each individual espouses (light for A and dark for B), and the weight of each connection represents the degree of trust the agents give to each other's evidence. As you can see, there are two groups with opposing beliefs who do not listen to each other.

We also find that the greater the distrust between those with different beliefs, the larger the fraction of the scientific community

that eventually ends up with false beliefs. This happens because those who are skeptical of the better theory are precisely those who do not trust those who test it. As this skepticism increases, more agents will fail to update their beliefs in light of new studies pointing toward more accurate beliefs. We can think of this sort of polarization as a way that communication closes down between opposed groups over time. The group holding false beliefs thus becomes insensitive to results pointing to better ones.

These results follow from one way of thinking about how scientists might distrust each other. But there are other possibilities. In a less dramatic version of the model, scientists would listen to everyone but discount the evidence of those who disagree with them rather than ignoring it completely. In models with this assumption, we find that all scientific communities eventually do reach a consensus, just as in the original Bala-Goyal models. But mutual mistrust slows the process dramatically. Even in cases where scientists listen to each other enough that they do not reach stable, polarized outcomes, mistrust among those with different beliefs can produce *transient* polarization—long periods during which some scientists prefer the worse theory and mostly discount the evidence of those who prefer the better one.

One of the more surprising aspects of this transient polarization is that people who start off holding similar positions can end up on opposite sides of a debate. Imagine, for example, that Sally and Joe are scientists, and Sally is initially a bit more skeptical than Joe about a new theory. If Joe gathers evidence supporting the theory, his credence will increase. Sally's credence will also go up, but not as much, because she trusts Joe's data less than Joe does. This means that both their credences are higher than before, but also farther apart. Now Joe gathers more evidence, and his beliefs again jump up. Sally is also more convinced, but since the distance between her

and Joe is even greater now, her credence changes even less than the first time.

Eventually, Sally may conclude that Joe's theory is better, but she will take a long time to get there. Or else Joe may approach certainty so much faster than Sally that he leaves her behind. From her perspective, it will look like he is going down a rabbit hole, and she will conclude that he is too radical to trust.

Of course, even transient polarization can be a damaging outcome. In cases like Lyme disease, dire consequences are associated with the wrong belief: either overtreatment with antibiotics or ignoring a dangerous chronic infection. A significant slowing of the emergence of scientific consensus can seriously affect the lives of those with the disease.

Polarization has been studied in many disciplines. There is a large literature, for instance, looking for explanations of polarization in individual psychology. But researchers in this field tend to assume that when two actors look at the same evidence, if they fail to change their beliefs in the same way, then at least one of them must be irrational.[58] After all, you might think, the evidence either supports a given belief or it does not.

For example, many psychologists have shown that people tend to search out and pay attention to only the evidence that accords with their current beliefs. This is known as "confirmation bias"—reasoning by which we tend to confirm our current beliefs—and it is a variety of what is sometimes called "motivated reasoning." A typical psychological experiment on polarization might give participants two sets of evidence, or arguments, for and against an issue, and see how they change their beliefs. Political scientists Charles Taber, Damon Cann, and Simona Kucsova, for example, presented

subjects with conflicting evidence on issues ranging from the legalization of marijuana to the Electoral College. They found that those who started with strong beliefs about these issues became only more entrenched during the study—irrespective of what their starting beliefs were or what evidence they were given.[59] The proposed explanation is that the subjects paid attention only to evidence supporting the view they already held.

We are not suggesting that this psychological effect does not occur. It seems it does—and it is very likely a factor in real-world polarization. But the models of polarization based on Jeffrey's rule that we have described strongly suggest that psychological biases are not necessary for polarization to result. Notice that our agents do not engage in confirmation bias at all—they update on any evidence that comes from a trusted source. Even if people behave very reasonably upon receiving evidence from their peers, they can still end up at odds.

These models can inform our understanding of political polarization as well as the polarization of a scientific group. Sometimes, polarization happens over a moral/social position. The abortion debate, for instance, is obviously extremely contentious, and most of the debate is not over facts but over whether it is inexcusably wrong to abort unwanted fetuses.

But in other cases, we see political polarization arise over matters of scientific fact. When it comes to climate change, for instance, the debate is not primarily about whether something is morally right or wrong, or whether an economic policy is just or not. Rather, the disagreement seems to be about whether carbon emissions from human sources actually contribute to changes in weather patterns. This is not a matter of morality or values: either greenhouse gases are affecting the climate, or they are not.

Of course, there is little question that industrial interests have obscured the scientific consensus on the causes of climate change,

by spreading misinformation and creating a sense of controversy. But the models we have discussed suggest that even without industrial interference, a community of people trying to choose scientific beliefs to guide their votes or policy choices can end up with this sort of disagreement.

The take-away is that if we want to develop successful scientific theories to help us anticipate the consequences of our choices, mistrusting those with different beliefs is toxic. It can create polarized camps that fail to listen to the real, trustworthy evidence coming from the opposite side. In general, it means that a smaller proportion of the community ultimately arrives at true beliefs.

Of course, the opposite can also happen: sometimes, too much trust can lead you astray, especially when agents in a community have strong incentives to convince you of a particular view. The models we have considered so far assume that all scientists accurately report their results. In this sort of case, it makes little sense to discount the results of those you disagree with. But this is not the universal case. In fact, in the next chapter, drawing on the modeling work of philosopher Bennett Holman at Yonsei University and philosopher and political scientist Justin Bruner at the Australian National University, we discuss how important discounting the evidence of others can be when industry attempts to influence science.

Ultimately, as we will see, when assessing evidence from others, it is best to judge it on its own merits, rather than on the beliefs of those who present it.

In 1846 Ignaz Semmelweis, a Hungarian physician, took a post in the first obstetrical clinic of the Vienna General Hospital. He soon noticed a troubling pattern. The hospital's two clinics provided free care for poor women if they were willing to be treated by students—doctors in the first clinic, where Semmelweis was stationed, and